CS 134: Networks — **Problem Set 1 Student Solutions**

Prof. Yaron Singer — Due: **Noon, Wednesday 2/1/2017**

Resources: This problem set contains 5 parts. All the relevant concepts can be found in the lecture and section notes. For further reading you may wish to consult Chapter 2 of Matt Jackson's book.

1. Diameter and Average Path Length (Jackson 2.2.2).

- The *distance* between two nodes is the length of a geodesic between them – i.e. the smallest number of links in the graph needed to connect the two nodes. For example, the distance between the nodes b and e in network A in the figure at the bottom is 2. If there is no path between two nodes, we say that the distance between them is infinite.
- The *diameter* of a network is the largest distance between any two nodes in the network.
- The *average path length* between the nodes of a network is the arithmetic mean of the distances between all pairs of nodes in the network. For example, consider the following network:

  $$(\{a,b,c\},\{ab,ac\}).$$

  The average path length is:

  $$\frac{1}{3}\Big[\text{distance}(a,b)+\text{distance}(a,c)+\text{distance}(b,c)\Big]=\frac{1}{3}\Big[1+2+1\Big]=\frac{4}{3}.$$

a. What is the distance between nodes e and i in network B?

b. What is the diameter of the following network?

$$(\{a,b,c,d,e\},\{ab,ac,ad,de\})$$

c. What is the average path length in the following network?

$$(\{a,b,c,d\},\{ab,ac,ad\})$$

Solution: From Victoria Basedow:

a. The distance between nodes e and i in network B is 5

b. The diameter of the following network is 3

$$(\{a,b,c,d,e\},\{ab,ac,ad,de\})$$

c. The average path length in the following network

$$(\{a,b,c,d\},\{ab,ac,ad\})$$

is as follows:

$$\frac{1}{6}\Big[\text{distance}(a,b)+\text{distance}(a,c)+\text{distance}(a,d)+\text{distance}(b,c)+\text{distance}(b,d)+\text{distance}(c,d)\Big]$$

$$=\frac{1}{6}\Big[1+1+1+2+2+2\Big]=\frac{3}{2}=1.5.$$

2. Cliquishness, Cohesiveness and Clustering (Jackson 2.2.3). Note we use the words "network" and "graph" interchangeably.

- A *complete subgraph* of a graph is a nonempty subgraph C with every two nodes in C linked. A *clique* of graph G is[1] a maximal complete subgraph of G. In other words, it is a complete subgraph C of G that is not a subgraph of any larger complete subgraph of G. For example, $(\{a, c\}, \{ac\})$ is a clique of network A but the following subgraph is not:

$$(\{b, d, e, f\}, \{bd, de, df, ef\})$$

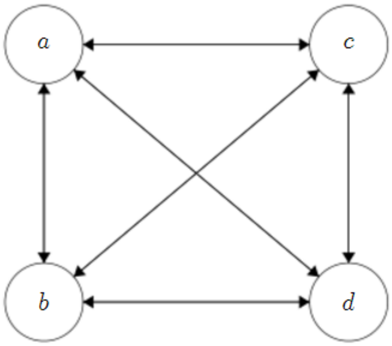
a. Represent a clique of network B via a picture.

- The most common way of measuring the cohesiveness of a network is based on transitive triples (or clustering). The most basic clustering measure is to perform the following exercise. Look at all situations in which two edges emanate from the same node (for example, edges ij and ik both involve node i) and ask how often the edge jk is then also in the network. We call this likelihood the *clustering coefficient*. For example, the clustering coefficient of network A is $\frac{3}{5}$. To see this, note that there are five pairs of nodes that are linked to a common node (bf, be, ef, df, de), and only three of these (de, ef, df) are themselves linked. In a friendship network, the clustering coefficient tells us how likely it is that two individuals with a common friend are friends themselves.[2]

b. Compute the clustering coefficient of network B.

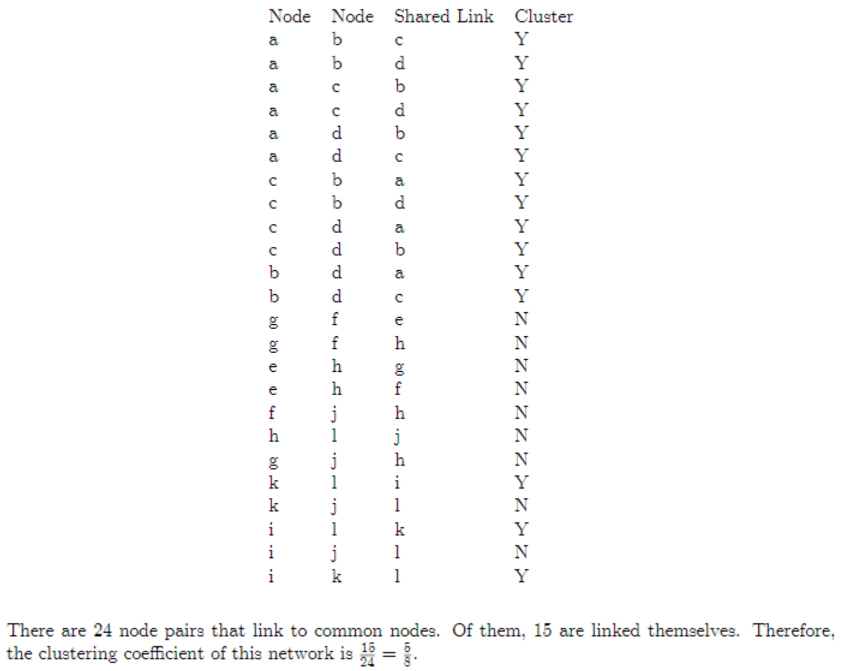
Solution: Solutions courtesy of David Choi.

a.



[1] Intuitively, cliques are the biggest completely connected pieces of a network.

[2] Another measure often used is the *average clustering coefficient*. To compute that, we look at each node i and compute i's clustering coefficient c_i—defined as the clustering coefficient of the subgraph consisting of i, its neighbors, and all links among them in the original graph. Then take the arithmetic mean of c_i across nodes. See Jackson for details.



Node	Node	Shared Link	Cluster
a	b	c	Y
a	b	d	Y
a	c	b	Y
a	c	d	Y
a	d	b	Y
a	d	c	Y
c	b	a	Y
c	b	d	Y
c	d	a	Y
c	d	b	Y
b	d	a	Y
b	d	c	Y
g	f	e	N
g	f	h	N
e	h	g	N
e	h	f	N
f	j	h	N
h	l	j	N
g	j	h	N
k	l	i	Y
k	j	l	N
i	l	k	Y
i	j	l	N
i	k	l	Y

There are 24 node pairs that link to common nodes. Of them, 15 are linked themselves. Therefore, the clustering coefficient of this network is $\frac{15}{24} = \frac{5}{8}$.

b.

3. Graph Cuts and Expansion.

- Given a graph $G = (V, E)$, and two disjoint subsets of nodes $S, T \subset V$ with $S \cap T = \emptyset$, a *cut* $C(S, T)$ is the set of all edges that have one end point in S and another endpoint in T. For example, if we have a graph $G = (V, E)$ with set of nodes $\{u, v, w, x, y, z\}$ and the set of edges $\{uw, ux, vx, vz, wx, xy, yz\}$, for $S = u, v$ and $T = w, x, y$ the cut is $C(S, T) = \{uw, ux, vx\}$.

- The *expansion* of a graph $G = (V, E)$ is the minimum, over all cuts we can make of the number of edges crossing the cut divided by the number of vertices in the smaller half of the cut. Formally:

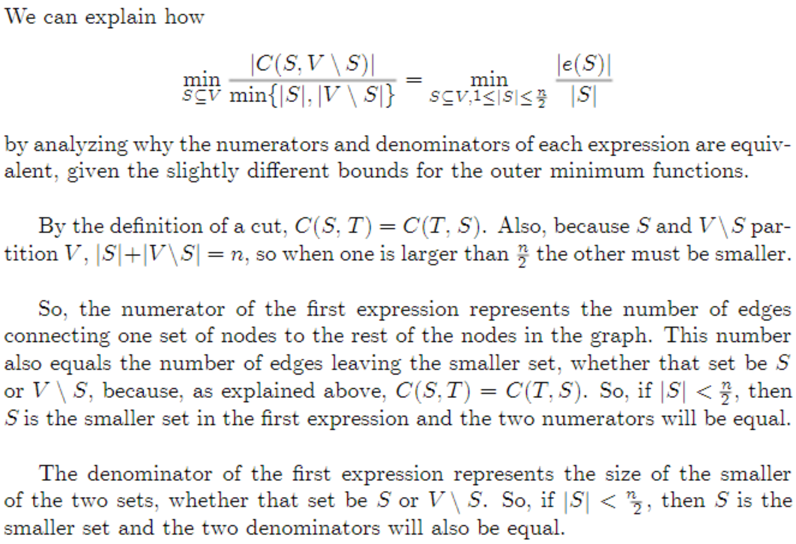
$$\min_{S \subseteq V} \frac{|C(S, V \setminus S)|}{\min\{|S|, |V \setminus S|\}}$$

Show that the expansion of a graph can also be expressed as α, where:

$$\alpha = \min_{\substack{S \subseteq V, \\ 1 \leq |S| \leq \frac{n}{2}}} \frac{|e(S)|}{|S|}$$

where $e(S)$ is the number of edges leaving the set S.

Solution: Solution courtesy of Matt Goodman.



We can explain how

$$\min_{S\subseteq V}\frac{|C(S,V\setminus S)|}{\min\{|S|,|V\setminus S|\}}=\min_{S\subseteq V,1\leq|S|\leq\frac{n}{2}}\frac{|\epsilon(S)|}{|S|}$$

by analyzing why the numerators and denominators of each expression are equivalent, given the slightly different bounds for the outer minimum functions.

By the definition of a cut, $C(S,T)=C(T,S)$. Also, because S and $V\setminus S$ partition V, $|S|+|V\setminus S|=n$, so when one is larger than $\frac{n}{2}$ the other must be smaller.

So, the numerator of the first expression represents the number of edges connecting one set of nodes to the rest of the nodes in the graph. This number also equals the number of edges leaving the smaller set, whether that set be S or $V\setminus S$, because, as explained above, $C(S,T)=C(T,S)$. So, if $|S|<\frac{n}{2}$, then S is the smaller set in the first expression and the two numerators will be equal.

The denominator of the first expression represents the size of the smaller of the two sets, whether that set be S or $V\setminus S$. So, if $|S|<\frac{n}{2}$, then S is the smaller set and the two denominators will also be equal.

4. Basics of Erdős-Rényi: Playing with Small Graphs. Consider the set of $n=4$ nodes $V=\{a,b,c,d\}$. Suppose that an edge (i.e., link) between any two nodes i and j forms with probability $p\in(0,1)$, independently of all other edges. This is the $G(n,p)$ Erdős-Rényi random graph with $n=4$.

a. For this graph on $n=4$ nodes, what is the probability that the graph has (i) no edges, (ii) one edge and (iii) more than one edge? The only variable in your answer should be p.

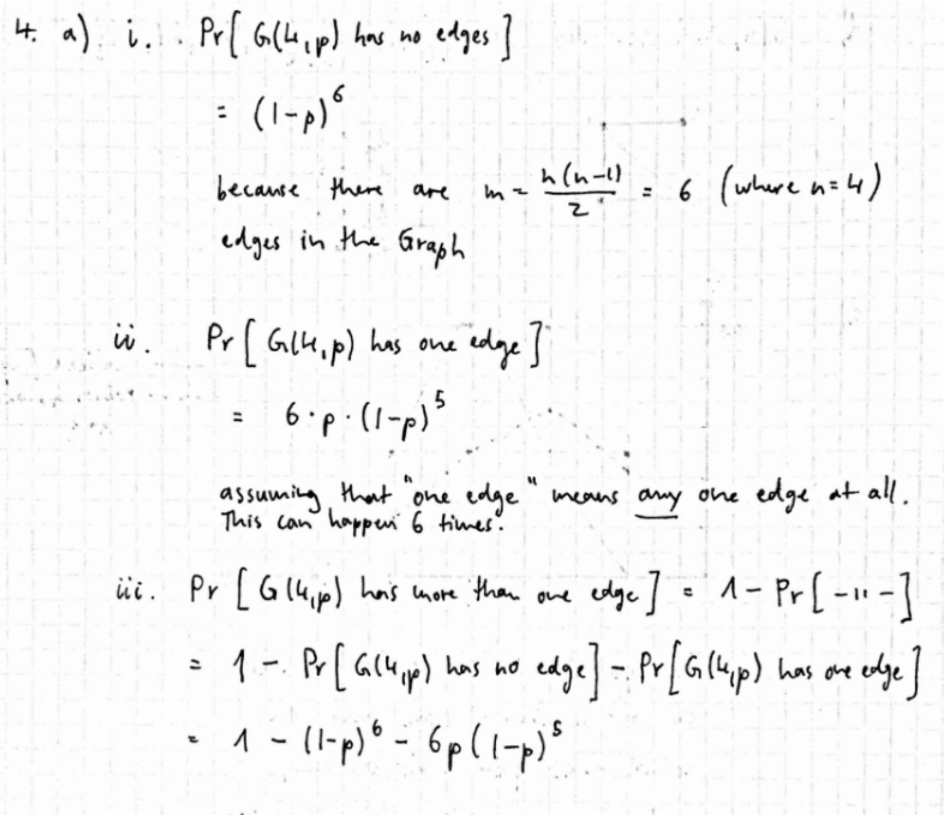
b. What is the expected degree of each node in $G(4,0.5)$?

c. Generate a random graph $G(4,0.5)$ on the nodes $V=\{a,b,c,d\}$ using a coin (as was done in class). Write down your graph (any of the descriptions introduced in Problem Set 1 are okay).

d. **Bonus question (not mandatory):** Plot the degree distribution of the graph you just made. Plot on top of it the Poisson distribution (see lecture notes for more information) with μ, the mean, equal to the expected degree you computed in (b). Comment on the quality of the approximation.

Solution:

a. Solution courtesy of Lars Lorch:



4. a) i. $\Pr[G(4,p) \text{ has no edges}]$

$= (1-p)^6$

because there are $m = \frac{n(n-1)}{2} = 6$ (where $n=4$) edges in the Graph

ii. $\Pr[G(4,p) \text{ has one edge}]$

$= 6 \cdot p \cdot (1-p)^5$

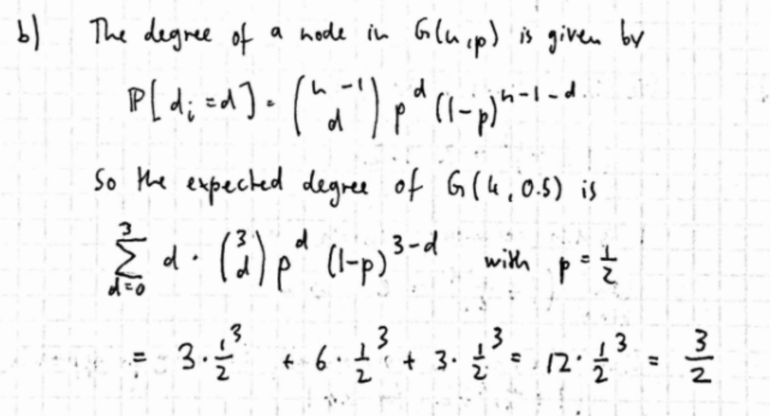
assuming that "one edge" means any one edge at all. This can happen 6 times.

iii. $\Pr[G(4,p) \text{ has more than one edge}] = 1 - \Pr[\,-\text{"}-\,]$

$= 1 - \Pr[G(4,p) \text{ has no edge}] - \Pr[G(4,p) \text{ has one edge}]$

$= 1 - (1-p)^6 - 6p(1-p)^5$

b. Solution courtesy of Lars Lorch:



b) The degree of a node in $G(n,p)$ is given by

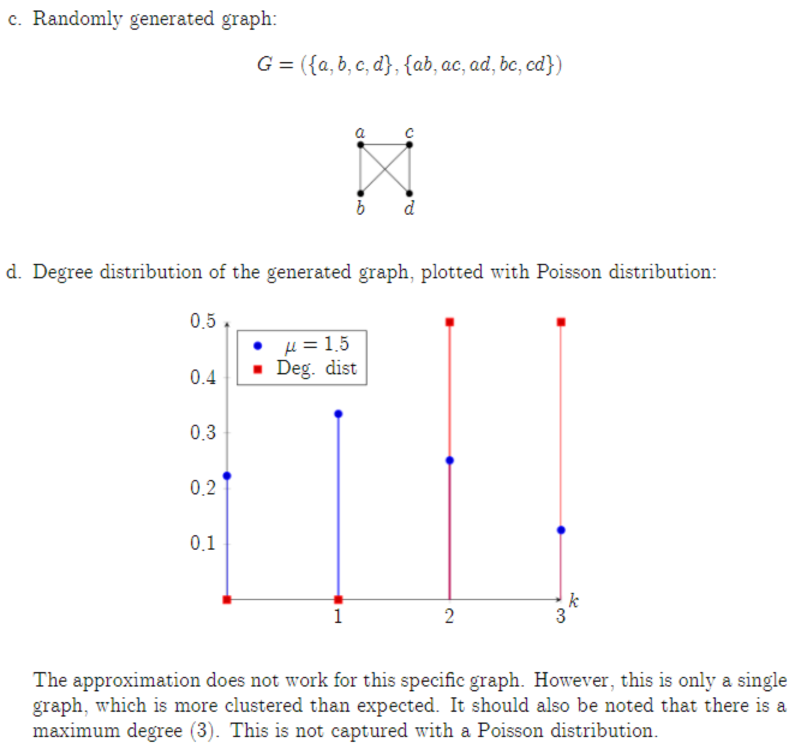
$$\mathbb{P}[d_i = d] = \binom{n-1}{d} p^d (1-p)^{n-1-d}$$

So the expected degree of $G(4, 0.5)$ is

$$\sum_{d=0}^{3} d \cdot \binom{3}{d} p^d (1-p)^{3-d} \quad \text{with } p = \frac{1}{2}$$

$$= 3 \cdot \frac{1}{2}^3 + 6 \cdot \frac{1}{2}^3 + 3 \cdot \frac{1}{2}^3 = 12 \cdot \frac{1}{2}^3 = \frac{3}{2}$$

c. Solution courtesy of James Kelleher:



c. Randomly generated graph:

$$G = (\{a, b, c, d\}, \{ab, ac, ad, bc, cd\})$$

d. Degree distribution of the generated graph, plotted with Poisson distribution:

The approximation does not work for this specific graph. However, this is only a single graph, which is more clustered than expected. It should also be noted that there is a maximum degree (3). This is not captured with a Poisson distribution.

d. See above.

5. Compound Growth. Fix an integer $k \geq 1$ and consider the sequence $s_n \equiv \left[(1+\frac{1}{k})^k\right]^n = \left[(1+\frac{1}{k})^k\right]^1, \left[(1+\frac{1}{k})^k\right]^2, \ldots$.

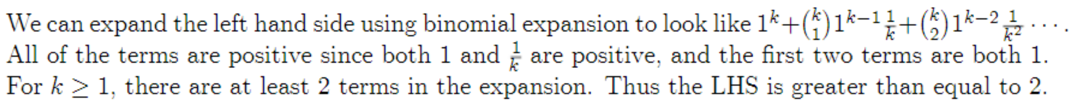
a. Show that $\left(1+\frac{1}{k}\right)^k \geq 2$.

b. Show that $n \geq 20$ implies $\left[(1+\frac{1}{k})^k\right]^n \geq 10^6$.

c. If you had one dollar to invest, and were guaranteed that you will earn p percent every day on your money, how long would it take you to become a millionaire (if you aren't one already)?

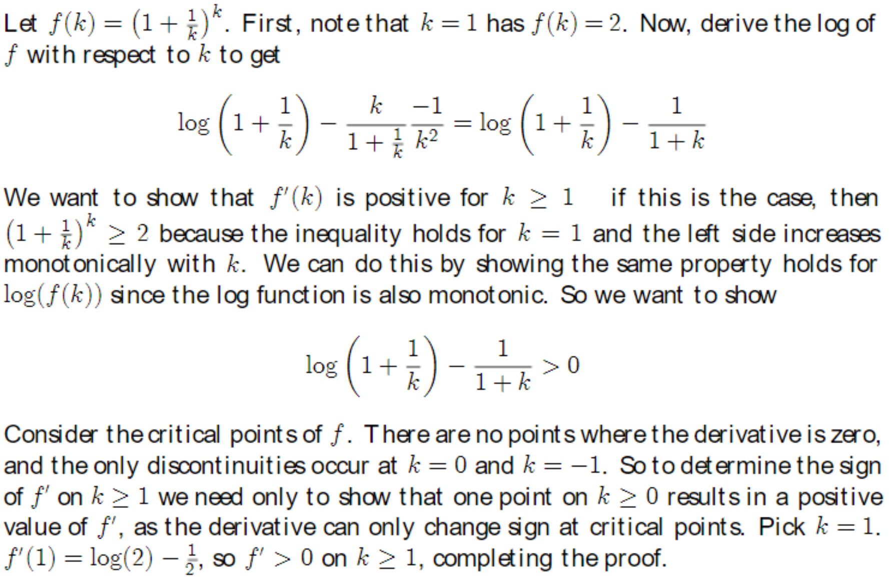
Solution:

a. Solution courtesy of Cecilia Zhou:



We can expand the left hand side using binomial expansion to look like $1^k + \binom{k}{1}1^{k-1}\frac{1}{k} + \binom{k}{2}1^{k-2}\frac{1}{k^2}\cdots$. All of the terms are positive since both 1 and $\frac{1}{k}$ are positive, and the first two terms are both 1. For $k \geq 1$, there are at least 2 terms in the expansion. Thus the LHS is greater than equal to 2.

Alternative solution courtesy of Matt Aguirre:



Let $f(k) = (1 + \frac{1}{k})^k$. First, note that $k = 1$ has $f(k) = 2$. Now, derive the log of f with respect to k to get

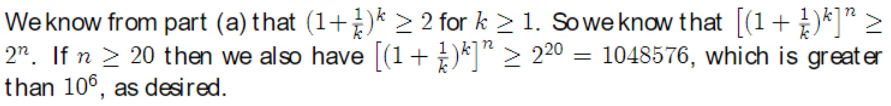
$$\log\left(1+\frac{1}{k}\right) - \frac{k}{1+\frac{1}{k}}\frac{-1}{k^2} = \log\left(1+\frac{1}{k}\right) - \frac{1}{1+k}$$

We want to show that $f'(k)$ is positive for $k \geq 1$ if this is the case, then $(1 + \frac{1}{k})^k \geq 2$ because the inequality holds for $k = 1$ and the left side increases monotonically with k. We can do this by showing the same property holds for $\log(f(k))$ since the log function is also monotonic. So we want to show

$$\log\left(1+\frac{1}{k}\right) - \frac{1}{1+k} > 0$$

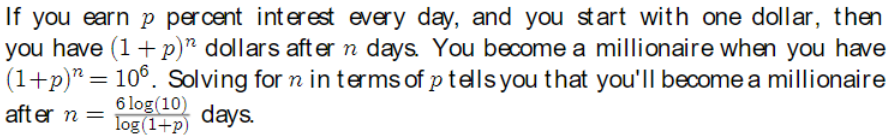
Consider the critical points of f. There are no points where the derivative is zero, and the only discontinuities occur at $k = 0$ and $k = -1$. So to determine the sign of f' on $k \geq 1$ we need only to show that one point on $k \geq 0$ results in a positive value of f', as the derivative can only change sign at critical points. Pick $k = 1$. $f'(1) = \log(2) - \frac{1}{2}$, so $f' > 0$ on $k \geq 1$, completing the proof.

b. Solution courtesy of Matt Aguirre:



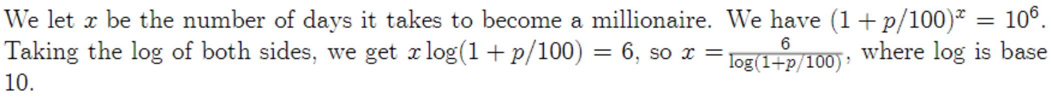
We know from part (a) that $(1+\frac{1}{k})^k \geq 2$ for $k \geq 1$. So we know that $\left[(1+\frac{1}{k})^k\right]^n \geq 2^n$. If $n \geq 20$ then we also have $\left[(1+\frac{1}{k})^k\right]^n \geq 2^{20} = 1048576$, which is greater than 10^6, as desired.

c. Solution courtesy of Matt Aguirre:



If you earn p percent interest every day, and you start with one dollar, then you have $(1+p)^n$ dollars after n days. You become a millionaire when you have $(1+p)^n = 10^6$. Solving for n in terms of p tells you that you'll become a millionaire after $n = \frac{6\log(10)}{\log(1+p)}$ days.

Alternative solution (different log base) courtesy of Cecilia Zhou:



We let x be the number of days it takes to become a millionaire. We have $(1 + p/100)^x = 10^6$. Taking the log of both sides, we get $x\log(1 + p/100) = 6$, so $x = \frac{6}{\log(1+p/100)}$, where log is base 10.

Figure 1: Network A

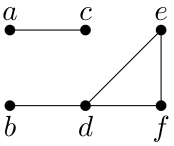
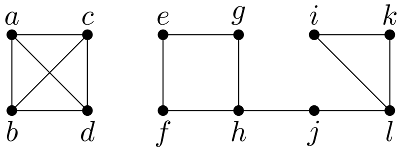


Figure 2: Network B


a
c
e
g
i
k
b
d
f
h
j
l